Automated Data Analysis for Geodetic Sensor Networks
Proposal for 632-07

Task Leaders

Andrea Donnellan, JPL, andrea.donnellan@jpl.nasa.gov, 818-354-4737
Jay Parker, JPL, jay.parker@jpl.nasa.gov, 818-354- 6790
John Rundle, U. Colorado, JPL Dist. Vis. Sci., rundle@hopfield.colorado.edu, 303-492-5642
Geoffrey Fox, Northeast Parallel Arch. Center, Syracuse U., gcf@npac.syr.edu, 315-443-2163

Product Description

NASA has identified the development of space-based and in situ sensor webs as a major technology requirement, especially by the Earth Science Enterprise. A crucial aspect of the development and deployment of sensor networks will be their ability to analyze data autonomously, so that the deluge of information provided by these devices can provide scientific insight, not solely enormous volumes of data. We propose to use space-based geodetic data as a testing ground to develop a prototype software and pattern recognition infrastructure, demonstrating the power of a wide-area network of data-collection assets when linked in a coordinated IT and analysis environment. This is a **new, push** task.

State-of-the-Art

		Improvement in Capability		
Component*	***Current***	***FY 2000***	***FY 2001***	***FY 2002***
1. Pattern recognition s/w	One aspect on a single fault	One aspect on 100's of faults	Many aspects, 100's of faults	Extension to other problems
2. Large datasets	Few tools	Object-oriented approach	Multiple datasets	Fully integrated with algorithms
3. Predictive capabilities	Coarse long-term forecasting	Months to produce	Weeks to produce	Continually updated

Benefits

NASA has recently committed a great deal of effort and funding to develop the means to observe and characterize the movements of the earth's crust that arise from plate tectonics and lead to catastrophic earthquakes. Understanding the dynamic processes responsible for these events will require not only a national commitment to develop the necessary observational datasets, but also the technology required to use these data in the development of sophisticated, state-of-the-art numerical simulations and models. The models can then be used to develop an analytical and predictive understanding of these large and damaging events, thus moving beyond the current, more descriptive approaches now routinely employed. This will allow hypothesis testing and data integration on a scale not previously possible.

The product will result in a dramatic increase in output compared to the current state of the art. At present researchers typically treat one aspect of the physics of an earthquake fault using one type of data. By using the methodology described below researchers will be able to integrate many aspects of hundreds of faults using multiple datasets into realistic models that describe the underlying physics of the system. The algorithms developed for the earthquake problem will be extensible to other problems in both space and earth science.

The object-oriented approach will result in standardized datasets and models. This will allow rapid and effortless sharing of information, retrieval of data, and model development and validation. Models will be physically consistent across a wide range of parameters and validated with real data. Only then will we be able to output realistic hazard maps that can be continually updated given the latest seismicity and surface deformation.

Understanding the patterns of systems may hold the key to understanding the physics of the systems themselves. We will reap enormous benefits by treating the earthquake problem by better understanding the behaviors of fault systems, but by also developing algorithms extensible to many other fields.

Technical Approach

There are five major objectives to the proposed task. 1) We will develop pattern recognition and datamining software using earthquake fault interactions as a testbed. 2) In order to do so we will develop the infrastructure to handle the heterogeneous large datasets and codes using an object broker oriented approach. 3) We will develop datamining algorithms for both large and sparse data sets. The earthquake problem is ideal for this because we have large volumes of data that are sparse in time because they sample a short part of the earthquake cycle. 4) We will address the NASA objective of developing predictive capabilities for natural hazards in general and, in particular, earthquakes. 5) We will produce general algorithms that are extensible to other sensor webs. Those involved in the Mars nav/comm constellation are enthusiastic about our potential involvement in their program.

We will develop spatio-temporal pattern recognition methods that uncover the most important correlations in geodetic time series data. These feature extraction techniques [1], that will ultimately enable autonomous monitoring systems, which can flag important events and patterns in real-time for further decision-making. Simultaneously, we will develop and deploy an associated modeling and collaboration framework that will enable analysis of the data by a number of remotely collaborating participants and integration with large-scale computational models and other datasets such as Interferometric Synthetic Aperture Radar (InSAR). Although demonstrated initially on geodetic data, the product developed here will emphasize general methods, using open software standards such as XML, to ensure its applicability across a range of future sensor network systems in ground-based, earth orbiting, and deep space environments. The effort will use data from the Southern California Integrated Global Positioning System (GPS) Network (SCIGN), a network of 250 GPS monitoring stations, seismicity data, and Interferometric Synthetic Aperture Radar (InSAR) data. Since our approach gains enormous leverage from an existing sensor network resource, we are in a magnificent position to focus on the crucial software issues of managing and analyzing data and models across a large distributed sensor array, without incurring the large overhead of device construction and deployment. The lessons learned on this existing 'testbed' can then be transferred at very little cost during the design phase of future sensor webs.

Earthquakes are part of a space-time system of interacting cells (fault segments) that are subjected to a persistent driving force. A fault segment fails when the force (or some physical variable for other cases) at a certain position and time reaches a predefined threshold. The faults interact with other faults by either bringing them closer to or inhibiting failure. The spatial and temporal patterns of failures or firings are richly complex and are difficult to understand from any deterministic point of view [2].

Recognizing and detecting space-time patterns in physical systems is frequently difficult, but in certain specific cases, considerable progress has recently been made [3–4]. If a driven nonlinear system has dynamics that are statistically stationary, the space-time patterns that appear are equivalent to the eigenvectors of certain classes of equal time correlation operators. The eigenvalues corresponding to these eigenvectors may represent probabilities, frequencies, slip amplitudes, or have other physical significance. Most of these correlation operators have not been extensively considered previously, so our results will have significant relevance to a variety of unexplored problems [1]. For example, in the driven systems we examine, one can form a time series that describes activity at discrete locations x at discrete times t using a time series y(x,t). If z(x,t) represents the zero-mean, univariant time series obtained from y(x,t), one can form a *state correlation operator:*

One can also form a *rate correlation operator*:

$$C(x,x') \equiv \frac{1}{T}\int_0^T dt\ z(x,t)\ z(x',t)$$

$$K(x,x') \equiv -\frac{1}{T}\int_0^T dt\ \frac{\partial z(x,t)}{\partial t}\ \frac{\partial z(x',t)}{\partial t}$$

The eigenvalues p_n of $C(x,x')$ represent probabilities at which the eigenvectors $\chi_n(x)$ are represented in the data, and the eigenvalues ω_n of $K(x,x')$ represent frequencies at which the eigenvectors $\phi_n(x)$ recur in the data. Both $\chi_n(x)$ and $\phi_n(x)$ represent patterns, $\chi_n(x)$ being spatial patterns, and $\phi_n(x)$ space-time patterns. Since the previously published methods have investigated only the basic correlation operators like C(x,x'), our work examining the implications of other types of correlation operators should break considerable new ground.

Status and Milestones

This is a new task. However, the PIs have been involved in certain aspects of this technology for several years. Donnellan is an expert in collecting, analyzing, and modeling geodetic data and has designed and deployed a network of autonomous GPS stations in the extreme environment of Antarctica [5–7]. Parker has experience in information technology, and electromagnetic modeling and has recently begun applying his expertise to earthquake related modeling problems [8–9]. Rundle has spent several years developing quasi-static dynamic codes of earthquake fault interactions and has recently begun applying pattern recognition to the problem of fault interactions and behavior [10–11]. Fox is at the forefront of the field of developing problem solving environments using object broker technologies and will enable the establishment of an infrastructure capable of handling the large volumes of data and complex models [12–14].

FY 2000 Milestones:

- Prototype time series algorithms implemented for a distributed array of geodetic sensors.
- Data handling and modeling environment design developed.

FY 2001 Milestones:

- Integrated feature extraction/modeling and visualization scenarios implemented.
- End-to-end prototype implemented for hypothesis evaluation/modeling/data integration.

FY 2002 Milestones:

- Testing and cross-comparison of space-geodetic data; computational models completed.
- Infrastructure deployed to demonstrate the scaling of autonomous data analysis/modeling to a 1000-node network such as the proposed Plate Boundary Observatory (PBO).

Customer Relevance

Our product will provide immediate benefits to the Plate Boundary Observatory (PBO) a proposed follow-on to SCIGN consisting of over 1000 sites across the western US. Geodetic missions such as Shuttle Radar Topography Mapper (SRTM) and LightSAR will provide unprecedented Digital Elevation Models over much of the world, which can also be incorporated into our framework. LightSAR will also provide valuable coseismic displacement fields and surface deformation information. More importantly, this work will point the way to an entire class of future sensor network applications for both the Space Science and Earth Science Enterprises by developing a systematic approach to the testing of scientific hypotheses in a distributed sensor environment characterized by both large data volumes and large modeling requirements.

One of the six major questions in the NASA Administrator's strategic outlook asks how we can develop predictive natural disaster models. These events (e.g. volcanic eruptions or earthquakes) are potentially catastrophic if not flagged quickly and reliably. The datamining and pattern recognition software to be developed in this program will be directly applicable to early warning systems. Our approach is based on the development of general pattern recognition and datamining. We also address NASA's basic questions: "What cutting-edge technologies, processes, techniques, and engineering capabilities must we develop to enable our research agenda in the most productive, economical, and timely manner? How can we most effectively transfer the knowledge we gain from our research and discoveries to commercial ventures in the air, in space, and on Earth?" By integrating these concepts within an object-oriented framework based on open standards, we can provide a blueprint for the design of fully autonomous continuously monitoring systems of heterogeneous assets that will underpin much of NASA/ESE's needs in the future.

Not only will this product apply to solid earth science in the study of the behavior of earthquake fault systems, but it will also exploit data from the regional and global GPS networks for estimation of water vapor content in the atmosphere and total electron count in the ionosphere. The evolving patterns of tropospheric water vapor and ionospheric electron content constitute evolving systems with highly detailed spatial and temporal correlations. These affect performance of NASA resources in a wide array of communication and remote sensing applications. Interpolation, assimilation and prediction are active research areas in both these fields that will benefit from this effort. Results from these studies impact our ability to communicate during ionospheric storms, and allow the removal of noise due to water vapor from InSAR measurements. Members of this team already work closely with people in the above mentioned fields.

The techniques developed here, then have direct application to the planned Mars nav/comm constellation (Mars network). This sparse constellation of satellites around Mars will be used to support Mars navigation; studies of crustal deformation on Mars; nutation, rotation, and polar motion measurements, which yield information about the presence of a liquid core; and atmospheric studies. The pattern recognition algorithms developed under this project will be crucial for navigation and datamining from this sparse Mars constellation.

Technical References

[1] Rundle, Klein, Tiampo, and Gross, “Linear Pattern Dynamics in Nonlinear Threshold Systems,” *Phys. Rev. E.*, manuscript submitted.

[2] Nijhout, H.F., L. Nadel, and D.L. Stein, *Pattern Formation in the Physical and Biological Sciences*, Lecture Notes V, Santa Fe Inst., Addison Wesley, Reading, MA, 1997.

[3] Fukunaga, K., *Statistical Pattern Recognition*, Academic Press, San Diego, 1990.

[4] Holmes, P., J.L. Lumley and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, 1996.

[5] Satellite Geodesy and Geodynamics Systems Group, GEM web page URL http://milhouse.jpl.nasa.gov/gem, GPS in Antarctica URL http://geodynamics.jpl.nasa.gov/antarctica

[6] Donnellan, A. and G. A. Lyzenga, Fault afterslip and upper crustal relaxation following the Northridge earthquake, *J. Geophys. Res.*, 103, 1998.

[7] Heflin, M.B., D. Dauger, D. Dong, A. Donnellan, K. Hurst, D. Jefferson, G. Lyzenga, M. Watkins, F. Webb, J. Zumberge, Rate change observed at JPLM after the Northridge earthquake, *Geophys. Res. Lett.*, **25**, 93–96, 1998.

[8] Cwik, T., C. Zuffada, D. Katz and J. Parker (1998),"Radar Scattering and Antenna Modeling on Scalable High-Performance Computers", in Industrial Strength Parallel Computing: Programming Massively Parallel Processing Systems, A. Koniges (ed.), Morgan-Kaufmann (in press).

[9] Parker, J.W., and S. A. Bowhill, “Mesospheric KHI Signatures in Reconstructed Broad-Band Rocket Probe Data,” J. Geophys. Res. 96, 1991.

[10] Rundle, J.B., W. Klein, S. Gross, and D.L. Turcotte, Boltzmann fluctuations in numerical simulations of nonequilibrium threshold systems, Phys. Rev. Lett., 75, 1658-1661, 1995.

[11] Rundle, J.B., W. Klein and S. Gross, Physical basis for statistical patterns in complex earthquake populations: Models, predictions and tests, Pure Appl. Geophys., in press, 1999.

[12] NPAC, Syracuse University, GEM web page URL http://www.npac.syr.edu/projects/gem/

[13] Fox, G.C.,Akarsu E., Furmanski W., Haupt T., "WebFlow -- High-level Programming environment and Visual Authoring Toolkit for High Performance Distributed Computing" in Proc. of SC98, 1998.

[14] Fox G.C., Furmanski W., "Computing on the Web,New Approaches to Parallel Processing,Petaop and Exaop Performance in the Year 2007 IEEE Internet Computing 1:2,38-46, 1997

Letters of Support

June 21, 1999

Dr. Andrea Donnellan
Satellite Geodesy and Geodynamics Systems Group
Mail Stop 238-600
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109-8099

Dear Andrea,

As we have discussed, I am keenly interested in your CETDP Program Proposal: "Automated Data Analysis for Geodetic Sensor Networks." Let me know if there is anything the Southern California Earthquake Center can do to help facilitate your efforts in this regard.

Space-based and in-situ sensor webs are important in earthquake science and engineering, and the ability to analyze massive amounts of data from these webs efficiently, autonomously, and in real time is sure to lead to greater scientific insights than has been possible in the past through routine data reduction and analysis. In particular, the vast new seismic and geodetic networks, both contemplated and currently under construction, will lend themselves to the kinds of pattern recognition techniques you propose to develop. Ultimately, one of the real beneficiaries of this type of research might be the area of improved seismic hazard assessment and concomitant decision making.

To date, only a fraction of the information generated by seismic and space-based crustal deformation networks has been used for understanding the earthquake process. The spatio-temporal pattern recognition and other data mining techniques you are proposing have the potential to uncover aspects of the earthquake cycle not yet recognized. The modeling and collaboration framework that are also part of your proposal should stimulate involvement by multiple groups of investigators, encourage multidisciplinary integration, and ensure wide applicability of your methodology.

In short, the technology you propose to develop will help pave the way for more fully exploiting sensor webs in many areas of earth and space science and engineering. It is generally believed that future progress in many areas of science will require increased resolution, which translates into more observations in space and time. In the earthquake business alone we are talking about factors of ten or more within the next decade! This, in turn, will require innovative approaches and a new generation of data handling and analysis techniques if we are to fully exploit the vast amount of new information. Your proposal directly addresses these needs. I hope you will keep me informed regarding its progress.

Sincerely,

Thomas L. Henyey

Professor of Geological Sciences, University of Southern California, and Director,
Southern California Earthquake Center

June 22, 1999

Dr. Andrea Donnellan
Satellite Geodesy and Geodynamics Systems Group
Mail Stop 238-600
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109-8099

Dear Andrea,

I am very interested in your CETDP Program Proposal: "Automated Data Analysis for Geodetic Sensor Networks." The geodynamics program within ESE has invested a tremendous amount in geodetic networks and techniques for studying earthquakes and volcanoes. The product that you describe in this proposal directly addresses our need for efficiently analyzing the great volume and variety of data now being collected. The tools that result will help provide an analytical and predictive understanding of large and damaging events from earthquakes and volcanoes, thus moving beyond the current, more descriptive approaches now routinely employed. Your product should help bring data collection and modeling efforts into a cohesive program within ESE's Natural Hazards Program.

To make the fullest use of the data from our missions and geodetic systems within ESE, a seamless data handling-/modeling-analysis environment must be established. This is the goal of your proposal. This goal would allow analysis by a greatly expanded number of science investigators, and also provide the tools for applications practitioners (such as disaster managers) to make use of such data. These two aspects — expanded science and applications — are central to our goals within the Earth Science Enterprise. I fully support your effort to bring this about through this proposal.

Sincerely yours,

Clark R. Wilson
Program Scientist for Geodynamics and Geopotential Fields
Code YO
NASA Headquarters

May 10, 2000

Dr. Andrea Donnellan
JPL 238-600
4800 Oak Grove Drive
Pasadena, CA 91109

Dear Dr. Donnellan:

I wish to encourage your proposed activity in "Automated Data Analysis for Geodetic Sensor Networks" under the UPN632 program. Although this proposed activity has a strong Earth science component, it would also have major and noticable benefits for the relatively new but fast growing Mars Network Project, which is under the auspices of the Mars Surveyor Program. The Mars Network Project, which is closely allied with the Mars Micromission Project, is implementing a series of flight missions to Mars to deploy an Infrastructure of multiple (half a dozen, maybe more) Mars orbiters (starting in 2003, followed by additional deployments in 2005, 2007, etc.) which will provide navigation and communication services to multiple missions in Mars orbit and on the surface of the planet. The concept emphasizes a network approach to the Infrastructure, with the orbiters, ground rovers, landers etc. implemented as nodes. The system resembles, in many ways, a sparse GPS-like system at Mars, with the added capability to support high-speed data links between the orbiters and users.

GPS at Earth started as a navigation system, but as a technology it rapidly evolved to revolutionize many different scientific and civilian disciplines. We envision that this will also happen at Mars. With the Nav/Comm Infrastructure deployment starting up, the development of sensor network technology will be very important to the success of this program. In addition, this Mars Infrastructure (just as with GPS at Earth) will enable many exciting new scientific measurement possibilities at Mars, including short and long term experiments to track atmospheric changes, crustal activity, Mars rotation, polar motion, and nutation. One specific experiment planned by the European community is called Netlander, and involves the deployment of a ground network at Mars to collect these types of data. Because the work you have proposed could contribute so directly to these future Mars science, I wish to encourage you to continue to pursue this work and to keep us informed about opportunities for joint development and collaborations across the different Enterprise boundaries.

Sincerely yours,

Stephen M. Lichten
Work Area Manager, Navigation and Radio Metrics
Telecommunications and Mission Operations Directorate, Technology Program
=========================
Jet Propulsion Laboratory
4800 Oak Grove Drive
Mail Stop 238-638A
Pasadena, California USA 91109